

## Reported emissions of organic gases are not consistent with observations

RONALD C. HENRY<sup>\*†</sup>, CLIFFORD H. SPIEGELMAN<sup>‡</sup>, JOHN F. COLLINS<sup>\*</sup>, AND EUNSUG PARK<sup>‡</sup>

<sup>\*</sup>Department of Civil & Environmental Engineering, University of Southern California, 3620 South Vermont Avenue, Los Angeles, CA 90089-2531; and  
<sup>‡</sup>Department of Statistics, Texas A&M University, College Station, TX 77843-3143

Communicated by Herbert Hauptman, Hauptman-Woodward Medical Research Institute, Inc., Buffalo, NY, April 21, 1997 (received for review August 5, 1996)

**ABSTRACT** Regulatory agencies and photochemical models of ozone rely on self-reported industrial emission rates of organic gases. Incorrect self-reported emissions can severely impact on air quality models and regulatory decisions. We compared self-reported emissions of organic gases in Houston, Texas, to measurements at a receptor site near the Houston ship channel, a major petrochemical complex. We analyzed hourly observations of total nonmethane organic carbon and 54 hydrocarbon compounds from C-2 to C-9 for the period June through November, 1993. We were able to demonstrate severe inconsistencies between reported emissions and major sources as derived from the data using a multivariate receptor model. The composition and the location of the sources as deduced from the data are not consistent with the reported industrial emissions. On the other hand, our observationally based methods did correctly identify the location and composition of a relatively small nearby chemical plant. This paper provides strong empirical evidence that regulatory agencies and photochemical models are making predictions based on inaccurate industrial emissions.

We have determined the make-up and impact of several industrial sources of organic gases in Houston, Texas, from ambient measurements and found these to be largely incompatible with self-reported emissions of organic gases in the area. These gases are man-made precursors to photochemical smog formation. Regulatory agencies and photochemical models rely on these self-reported industrial emission rates, which are often outdated, incomplete, or inaccurate. Our results provide an independent, objective estimate of industrial source compositions and contributions to organic gas pollution. Our approach uses measurements from an automated gas chromatography monitor at a site near the Houston Ship Channel (a large petrochemical complex) from June to November, 1993. Multivariate receptor modeling (1) was applied to hourly observations of total nonmethane organic carbon (TNMOC) and 54 hydrocarbon compounds from C-2 to C-9, as listed in Table 1. The impact of three sources determined by multivariate receptor modeling peaked strongly at certain wind directions, indicating that these should lie close to the monitoring site. However, we were only able to match the smallest of these sources to a specific industrial facility. The two largest sources were incompatible in direction and composition with emissions reported by industries in the vicinity of the monitoring site.

The data were obtained from the Texas Natural Resources Conservation Commission, which were responsible for running the monitoring site. The sampling and analysis procedures followed were those of the U.S. Environmental Protection Agency's Photochemical Assessment Monitoring Stations Pro-

gram. The period in question is June 18, 1993, to November 30, 1993. An automated gas chromatograph reported hourly average concentrations in ppb C of 54 volatile organic compounds (VOC) and TNMOC. TNMOC is the sum of all the peaks in the chromatogram, identified or not, and thus is always greater than the sum of the identified species. The identified species in Table 1 are all hydrocarbons; the system could not identify oxygenated or chlorinated organic compounds. The data were screened to remove outliers and missing data leaving a final data set of 2,541 hourly observations. The average TNMOC for this set was 545 ppb C.

Other data supplied with the organic gas data were hourly average ozone, carbon monoxide, nitrogen oxides, and wind data. The wind data consisted of hourly average wind direction, standard deviation of the wind direction (a measure of the variability of the wind direction), and resultant wind direction and speed (obtained by a vector average of the wind velocity). The wind sensor was mounted about 7 m above the ground, and 3 m above the nearest obstructions. We also obtained from the Texas Natural Resources Conservation Commission a detailed emission inventory of all the industrial sources of nitrogen oxides and organic gases in the Houston area. This included the description, location, and emission rates (in tons per year) of all the individual sources in a company facility. Emission rates are sometimes given for individual organic compounds, but often these are only given as unidentified VOC, or unidentified olefins, for example.

Source compositions and contributions to TNMOC were estimated by the source apportionment by factors with explicit restrictions (SAFER) multivariate receptor model (2). The SAFER model has been applied to a similar set of VOC data from another site in Atlanta, Georgia, which was dominated by vehicle-related emissions (3). The source compositions determined by SAFER for the Atlanta data closely matched source samples taken during the study; to this extent, the SAFER model can be considered validated. The average source contributions to TNMOC determined by SAFER are roadway, 18%; industrial 1, 20%; industrial 2, 12%; industrial 3, 5%; industrial 4, 5%; industrial 5, 17%; with 24% unidentified.

To help assure that we reached plausible hypotheses, we analyzed the hourly concentration data set by other methods. We independently modeled 12 important compounds (indicated in Table 1). These compounds are fit using least squares, to the model:  $C = \alpha P + \epsilon$ . The model also includes constraints that force the fitted compositions to be nonnegative. Here  $C$  denotes the measured hourly concentrations,  $\alpha$  represents the hourly impacts of each pollution source, and  $P$  represents the source compositions (profiles). The error is represented by  $\epsilon$ . Our model fits both the composition and impacts simultaneously.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/946596-4\$2.00/0

Abbreviations: TNMOC, total nonmethane organic carbon; VOC, volatile organic compounds; SAFER, source apportionment by factors with explicit restrictions.

<sup>†</sup>To whom reprint requests should be addressed.

Table 1. Composition of industrial sources (percent of TNMOC)

No.	Species	Industrial 1	Industrial 2	Industrial 3
1	Ethane*	2.08	0.68	1.46
2	Ethene	0.42	0.33	0.01
3	<i>n</i> -Propane*	2.74	-0.97	0.73
4	Propene	1.18	0.38	0.73
5	Isobutane	3.62	-0.03	-0.12
6	<i>n</i> -Butane	3.47	-0.07	-0.06
7	Acetylene*	0.01	0.02	0.01
8	<i>trans</i> -2-Butene	0.11	0.06	0.07
9	1-Butene+isobutylene	0.15	0.08	0.10
10	<i>cis</i> -2-Butene	1.91	0.41	-0.39
11	Isopentane+cyclopentane	18.96	0.30	1.13
12	<i>n</i> -Pentane	6.20	1.07	0.19
15	2-Methyl-2-butene	0.48	0.00	0.00
16	Cyclopentene	0.05	0.03	0.04
17	<i>trans</i> -2-Pentene*	4.52	0.00	0.00
18	3-Methyl-1-butene	5.88	0.00	0.01
19	1-Pentene	2.30	0.06	0.09
20	<i>cis</i> -2-Pentene	2.38	0.03	0.04
21	2,2-Dimethylbutane	0.44	0.19	0.27
22	2,3-Dimethylbutane*	1.16	2.94	-0.07
23	2-Methylpentane*	4.08	14.90	0.00
24	3-Methylpentane	2.25	15.13	0.42
25	Isoprene	0.10	0.05	0.06
26	4-Methyl-1-pentene	0.02	0.01	0.01
27	2-Methyl-1-pentene	0.01	0.01	0.01
28	<i>n</i> -Hexane	2.83	2.31	1.09
29	<i>trans</i> -2-Hexene	0.30	0.00	0.00
30	<i>cis</i> -2-Hexene	0.06	0.04	0.03
31	Methylcyclopentane	0.60	0.32	0.37
32	2,4-Dimethylpentane	0.25	0.16	0.15
33	Benzene	0.68	1.38	-0.06
34	Cyclohexane	0.13	0.07	0.08
35	2-Methylhexane	0.79	3.35	0.07
36	2,3-Dimethylpentane*	0.39	0.30	0.24
37	3-Methylhexane*	1.00	3.07	1.23
38	2,2,4-Trimethylpentane	0.42	0.11	0.05
39	<i>n</i> -Heptane	0.43	0.24	0.28
40	Methylcyclohexane	0.26	0.14	0.16
41	2,3,4-Trimethylpentane	0.12	0.07	0.08
42	Toluene	1.37	4.50	0.84
43	2-Methylheptane	0.23	0.09	0.13
44	3-Methylheptane	0.17	0.07	0.20
45	<i>n</i> -Octane	0.23	0.15	0.51
46	Ethylbenzene*	0.36	2.32	0.85
47	<i>m</i> -Xylene+ <i>p</i> -xylene*	0.19	-0.37	20.60
48	Styrene	0.05	0.03	0.03
49	<i>o</i> -Xylene	0.24	0.13	0.15
50	<i>n</i> -Nonane	0.09	0.05	0.06
51	Isopropylbenzene	0.03	0.02	0.02
52	<i>n</i> -Propylbenzene	0.05	0.03	0.03
53	1,3,5-Trimethylbenzene*	0.04	0.32	0.03
54	1,2,4-Trimethylbenzene*	0.25	0.92	-0.40
	Sum	76	55	32

\*Species used to develop the model.

If the composition is known, the estimated impact can be estimated by ordinary least squares. The fitting algorithm uses the least square formula for the impacts as a function of compositions—i.e., minimizing the sum of squared deviations over feasible source compositions. As with all nonlinear optimizations, the initial value that is used in the optimization is important. We used the SAFER results as starting values. While it is always possible that we found local minimums, graphs of the solutions look reasonable when plotted on the principal component axes. In addition we used sensitivity analysis with the algorithm's stopping criteria, so that to the

extent possible we assured ourselves that we found a good local minimum and hopefully a global minimum. The algorithm that we use comes from the MATLAB Optimization Toolbox computer package; it is called "constr."

For the three industrial sources described in this paper there is close agreement between this approach and the SAFER results. The  $R^2$  values between the two analysis methods were 80% for roadway, 98% for industrial source 1, 99% for industrial source 2, 99% for industrial source 3, 95% for industrial source 4, and 89% for industrial source 5. The averaged jackknifed uncertainty estimates based upon five samples each using 90% of the data for four major compounds of the industrial 1 source composition are less than  $\pm 3\%$ , relative standard deviation. The corresponding uncertainty estimate for industrial 2 is  $\pm 2.5\%$ , relative standard deviation. The relative error for meta/para-xylene for industrial source 3 was  $\pm 2\%$ . The percents are  $\pm 1$  relative standard error for the estimate of composition and the appropriate multiplier  $t$ -distribution percent point for 95% confidence intervals is 2.78. These uncertainties take into account imprecision of the measurements and estimates, but do not take into account systematic errors (biases) in the measurements or in the estimates. The main suspected sources of systematic error in the fitting procedure are the unmodeled emission sources and the possibility that a local minimum rather than a global minimum was found.

The relationship between TNMOC and the VOC sources in the emissions inventory is shown in Fig. 1. This figure is a map of the Houston ship channel area near the monitoring site, with two polar plot overlays that show the average nonvehicle TNMOC concentrations and the sum of the emissions in  $10^\circ$  sectors. The contribution of vehicles to TNMOC was calculated by using acetylene as a tracer with an abundance of 5.4%. Thus, nonvehicle TNMOC concentration for each hour is the TNMOC concentration less the acetylene concentration divided by 0.054. The average nonvehicle TNMOC was calculated for periods of relatively steady winds when the standard deviation of the winds was less than  $20^\circ$ . The emissions in the polar plot have been weighted by the inverse of the distance from the monitoring site before being summed over the  $10^\circ$  sectors. Thus, nearby emissions are given greater weight. Inverse distance weighting is equivalent to assuming ground level release of the pollutants and a moderately unstable atmosphere. For reference, Fig. 1 also shows the location of all VOC sources in the inventory that emit more than 10 tons of VOC per year and are within 5 miles of the monitoring site. This distance was chosen because there was a natural break in the sources at that distance and because 5 miles is less than 1 hr transport time for typical wind speeds.

From Fig. 1, we see that high TNMOC concentrations are associated with winds from the east, where there are many large refineries and other sources not shown on the figure at distances of 20–30 km. Winds from the east at this site are infrequent. Much more frequent, especially in the summer, are winds from the south and southeast. High TNMOC is associated with winds from the southeastern sector, where there are a number of large, nearby sources. The distance-weighted emissions show peaks to the east, southeast, and south-southeast. Of particular interest is the peak in emissions to the southeast since it seems to be almost coincident with a minimum in the nonvehicle TNMOC. In the following we will examine in greater detail the relationship between observations and specific VOC sources in the southeastern sector.

The amount of TNMOC attributed to the first three industrial sources increased strongly with wind from the south-southeast and the south, as seen in Fig. 2. This implies that these sources must lie close to the monitoring site in the direction given by the peak. Thus, it should be possible to identify these three sources with companies from the emissions inventory that lie in the same general direction. The compo-

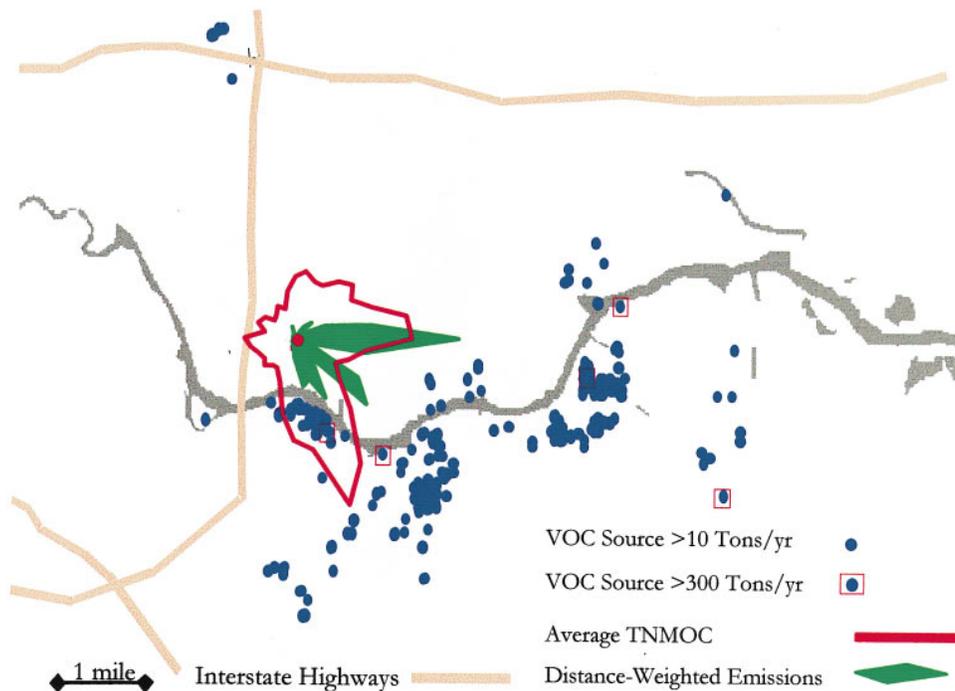


FIG. 1. Location of the monitoring site with respect to the Houston ship channel and nearby major VOC sources. Also shown are polar plots (in 10° sectors) of the average TNMOC concentration (red) and total VOC emissions (green). The TNMOC averages are for periods of steady winds—i.e., the standard deviation of the wind direction is less than 20°. The VOC emissions in the inventory have been weighted by the inverse distance from the site to give greater emphasis to nearby emissions. Only major VOC sources within 5 miles are shown, because this is typically less than 1 hr transport time.

sition of the sources determined by receptor modeling (given in Table 1) should also be consistent with the emissions inventory. Table 2 gives the emissions in tons of VOCs per year for 10° sectors from 130 to 210° relative to the monitoring site (all directions are azimuthal angles measured clockwise from north). This sector is especially of interest because it covers a large peak in TNMOC and 95% of the listed emissions are within about 4 km of the monitoring site.

We must compare Table 2 with Fig. 2. The amount of TNMOC from industrial 1 peaks at 160–170, and the amount from industrial 2 and 3 peaks at 180–190. So we expect to find large, nearby sources in these directions. In the 180–190 direction range, the largest source is the Bayer Corporation.

According to the emission inventory 51% of its emissions are xylenes with the remainder mostly chlorinated hydrocarbons. This is consistent with the composition of industrial 3 in Table 1. We therefore identify industrial 3 as emissions from the Bayer Corporation.

The impact of industrial 2 also peaks in the 180–190 direction and is larger than industrial 3, but there are no other large sources in this direction. The composition of industrial 2 is not consistent with the emissions inventory of the Bayer Corporation: industrial 2 has an unusually high amount of ethylbenzene. A search of the emissions inventory does not reveal any large source of ethylbenzene in this direction. Thus, we cannot unequivocally associate any source in the emission inventory with industrial 2.

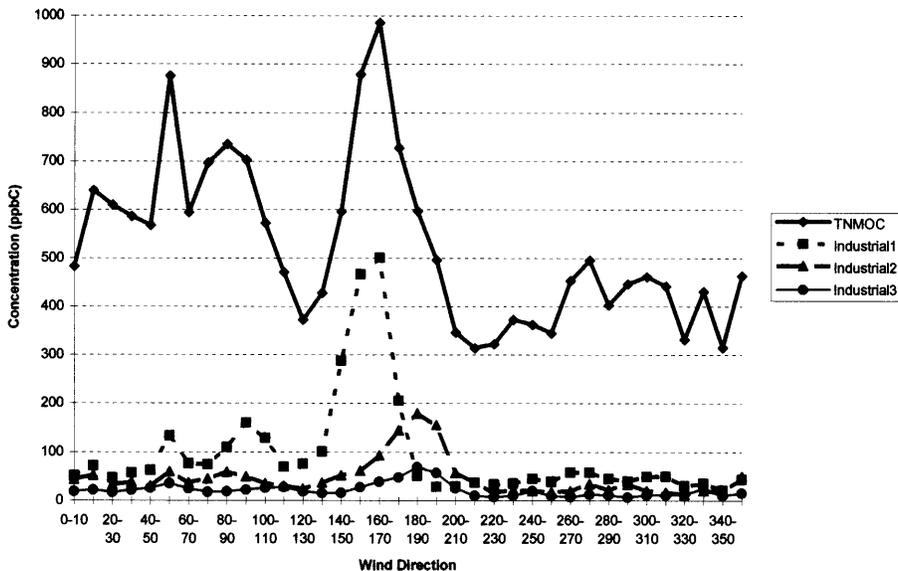


FIG. 2. Average concentration of TNMOC and three industrial sources for 10° sectors.

Table 2. Self-reported emission rates of organic gases (tons/year)

Company	Distance, km	Direction								Grand total
		130–140	140–150	150–160	160–170	170–180	180–190	190–200	200–210	
Phibro Energy USA, Inc.	1.4	0	4	1,141	639	281	121	94	55	2,338
Gatx Terminals Corporation	1.5	1	0	0	0	0	0	0	0	2
Shell Oil Company	2.9	0	1	0	0	0	0	0	0	4
Goodyear Tire and Rubber Co.	3.3	0	0	0	0	134	10	0	0	147
Lyondell CITGO Refining Co.	3.4	3,026	863	78	54	0	0	0	0	4,024
Mobil Chemical Company	3.4	0	0	0	109	2	0	0	1	116
Bayer Corporation	3.6	0	0	0	0	0	434	0	0	438
Texas Petrochemicals Corp.	3.9	352	0	0	10	1,082	9	0	0	1,457
Gulf States Asphalt Company, Inc.	8.1	0	0	0	136	0	0	0	0	144
Exxon Corporation	17.5	133	0	0	0	0	0	0	0	151
NASA	25.2	8	63	0	0	0	0	0	0	96
Grand total	74.3	3,520	931	1,220	948	1,500	574	94	57	8,917

This is also the case for industrial 1, the largest single source determined by receptor modeling. When the winds blow from the 150–170 sector, it accounts for about half of the TNMOC on average, and over 80% for the times of maximum impact. In this sector, the Phibro Energy refinery accounts for 82% of the listed emissions and is only 1 km from the monitoring site. It is reasonable to associate this refinery's emissions with industrial 1. However, the composition of industrial 1 is very high in C-5 olefins, and does not resemble gasoline or gasoline vapor, which might be expected for emissions from a refinery. Unfortunately, 83% of the listed emissions of Phibro are given as unidentified VOC, so we can only speculate on the likely composition of emissions from this source. The only major source of olefins of any kind in the inventory is the Texas Petrochemicals Corporation (42% butene and unidentified olefins), but these emissions are located at an angle of about 178 from the monitoring site, and thus are in the wrong direction. We conclude that industrial 1 cannot be positively identified with any source in the emissions inventory.

We were able to match only one source from the receptor modeling results with a specific chemical plant in the emissions inventory. The largest source in Table 2, the Lyondell CITGO refinery does not fulfill our expectations for the direction or size of its impact. It lies in a direction (130–140) that is not consistent with either our industrial 1 source or the observed TNMOC. Furthermore, its impact should be greater than 481 ppb C. This estimate follows from the ratio of the peak height of industrial 3 (Bayer Corp.) to its emissions 69/434, times 3026, the emissions of the Lyondell refinery in the direction 130–140. This is an underestimate of the impact of Lyondell because it is closer to the site than the Bayer Corp. Thus, the emission inventory fails to agree with our observation-based approach in two regards: it fails to account for sources that we

see, and large sources in the inventory do not show up in the observations. One could question our methods, except we are able to identify a rather small source, the Bayer Corp. Even if our results are discounted, it is impossible to reconcile the peak in TNMOC at 160–170 with the emissions inventory, which shows large emissions from the 130–140 sector. We believe that our results show that the inventory of industrial VOC emissions is inaccurate in its location, composition, and emission rates of major sources. In spite of the best efforts of government and industry, the emissions from refineries and chemical plants are notoriously hard to determine. Most of the emissions are so-called fugitive emissions from leaking valves, pipes, or connectors, of which there are tens of thousands in a large facility. So the failure of the emissions inventory to compare with the observation-based results is perhaps not surprising. One possible solution to this problem are methods that estimate emission rates based on observations (4).

We thank the Texas Natural Resource Conservation Commission for supplying the data and partially supporting this study, and Mr. Gene McMullen of the Houston Air Pollution Control for additional wind data. C.H.S. and E.P. were supported by Grant DMS-9523878 from the Chemistry and Statistics and Probability programs at the National Science Foundation. R.C.H. and J.F.C. were partially supported by Cooperative Agreement CR822072 with the U.S. Environmental Protection Agency.

1. Henry, R. C. (1991) in *Receptor Models for Air Quality Management*, ed. Hopke, P. K. (Elsevier, Amsterdam), pp. 117–148.
2. Henry, R. C. & Kim, B. M. (1990) *Chemom. Intell. Lab. Syst.* **8**, 205–216.
3. Henry, R. C., Lewis, C. W., & Collins, J. F. (1994) *Environ. Sci. Technol.* **28**, 823–832.
4. Henry, R. C., Wang, Yi-Jin, & Gebhart, Kristi A. (1991) *Atmos. Environ. Part A* **25A**, 503–509.